Preprint
UCRL-JC-136053

Multiple Input Feature Sets from Real-Time Color and Range Data for Reliable Tracking

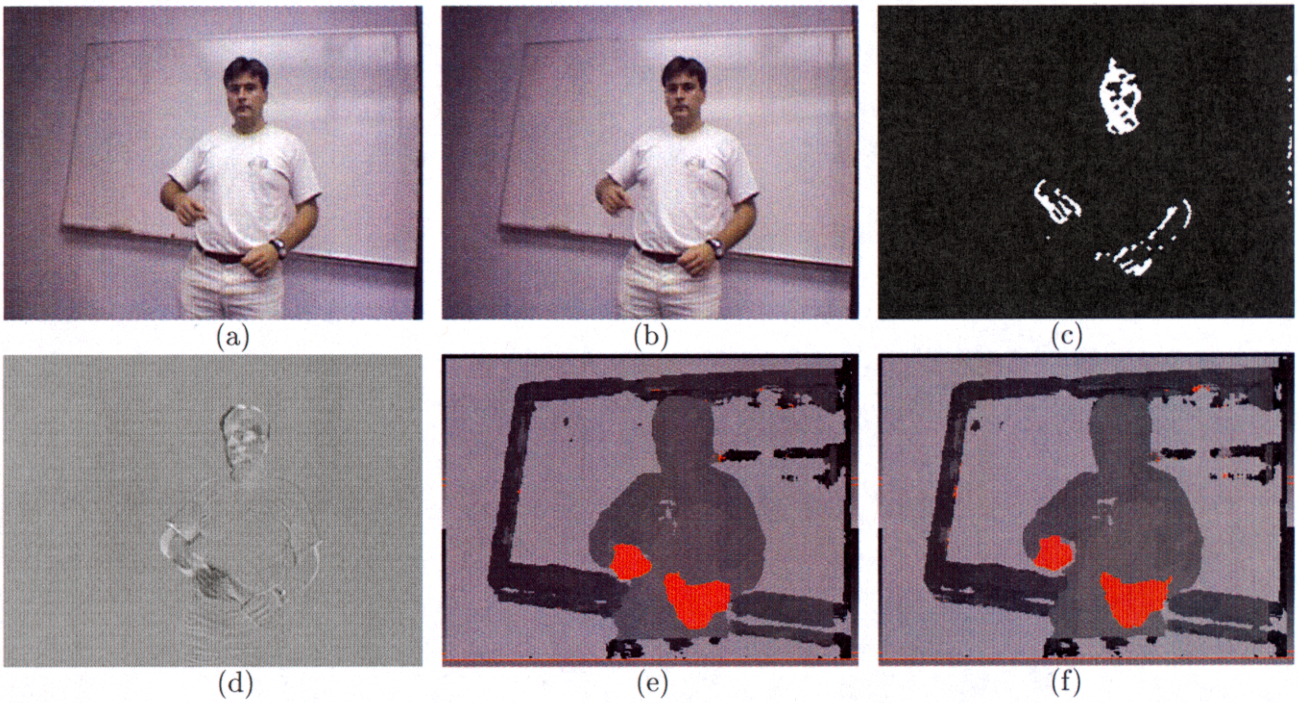
L.V. Tsap and D.B. Goldgof

This article was submitted to
Workshop on Digital and Computational Video (DCV '99)
Tampa, FL
December 10, 1999

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

October 19, 1999



(a) (b) (c)

(d) (e) (f)

Figure 1: (a-b) Intensity images of the speaker. (c) Results of skin segmentation and thresholding. (d) Image subtraction. (e-f) Range images with selected depth levels.

fully tested for the face and gesture tracking application. Search in the range domain prevented us from considering intensity-based segmentation artifacts (such as a curtain and a belt). Segmentations along intensity and depth channels can be also done independently and then combined. Future work will also include gesture recognition in 3-D and real time for specific interfaces.

Acknowledgments

We would like to thank Vladimir Tucakov at the Point Grey Research for providing the image sequences.

References

[1] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on PAMI*, 19(7):780–785, July 1997.

[2] M.-H. Yang and N. Ahuja. Recognizing hand gestures using motion trajectories. In *Proceedings of IEEE CS Conference on Computer Vision and Pattern Recognition*, volume 1, pages 466–472, Fort Collins, CO, June 1999.

[3] J. M. Rehg and T. Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. *Proc. of European Conference on Computer Vision*, 2:35–46, May 1994.

[4] J. J. Kuch and T.S. Huang. Model-based tracking of self-occluding articulated objects. In *Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration*, pages 666–671, MA, June 1995.

[5] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 90–102, San Juan, Puerto Rico, June 1997.

[6] L. V. Tsap, D. B. Goldgof, and S. Sarkar. Model-based nonrigid motion recovery from sequences of range images without point correspondences. In *Proceedings of IEEE Signal Processing Society 1998 International Conference on Image Processing ICIP98*, volume 2, pages 218–222, Chicago, IL, October 1998.

Multiple Input Feature Sets from Real-Time Color and Range Data for Reliable Tracking *

Leonid V. Tsap

Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
P. O. Box 808, L-551, Livermore, CA 94551
tsap1@llnl.gov

Dmitry B. Goldgof

Department of Computer Science and Engineering
University of South Florida
Tampa, Florida 33620
goldgof@csee.usf.edu

Abstract

This paper describes a work in progress on using multiple sets of input features for robust real-time object tracking in image sequences. Traditional approaches to tracking relied mostly on segmentation of the intensity data using motion or appearance data. Recent availability of real-time range data allows us to use it as an additional unrivaled source of information. We propose a combination of intensity- and range-based input features. Range data enables localized search for specific features which improves tracking reliability and speed. Proposed approach was successfully tested for the face and gesture tracking application.

1 Introduction

Object tracking from image sequences is a very important research domain. Goals of object tracking include segmentation of each frame into differently moving objects, selecting the object of interest and analyzing its motion during the entire sequence or multiple sequences. Therefore, object tracking involves processing of both spatial and temporal data. A number of applications are dealing with tracking motion of human bodies. These applications include video-surveillance, gesture-based interfaces to multimedia applications and systems, interfaces for people with disabilities preventing them from using the standard input technology, and videoconferencing. There is a continuing effort to involve human communication

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract number W-7405-Eng-48. UCRL-JC-XXXXXX

through movement in the design and development of computer interfaces that adequately capture such natural forms of communication. Another application is object manipulation in virtual environments.

Traditional approaches to tracking relied mostly on segmentation of the intensity data using motion or appearance data. Majority of the methods starts out by segmenting the human body from the background. For instance, in blob approaches people are modeled by blobs resulting from pixel classification based on color and position in the image. Wren *et al.* [1] achieved segmentation by classifying pixels into one of several models, including a static world and a dynamic user represented by gaussian blobs. Yang and N. Ahuja [2] used skin color and geometry of palm and face regions for segmentation stage of their system. A Gaussian mixture (with parameters estimated by an EM algorithm) modeled distribution of skin color pixels. Regh and Kanade [3] used a 3-D hand model to track a hand. They compared line features from the images with the projected model and performed incremental state corrections. Similar work was presented by Kuch and Huang [4] where the synthesis process could fit the hand model to any person's hand. A review by Aggarwal and Cai [5] classified approaches to the human motion analysis, the tasks involved, and major areas related to human motion interpretation.

2 Description of the Method

Registered range data provides an additional information valuable for segmentation and tracking. Often, an object of interest can be separated from other objects or background by depth alone. In other cases, having less artifacts (that could complicate segmentation) in range information comparing to intensity data is an important consideration [6]. However, until few months ago the solution was considered slow and expensive.

Recent availability of real-time registered range data (in addition to a color intensity information) allows us to use it as an additional unrivaled source of information. We propose a combination of intensity- and range-based input features. Range data enables localized search for specific features which improves tracking reliability and speed.

Real-time constraints such as temporal correlation produces a possibility of searching within a smaller region based on the match in the previous frame. For the range image this involves depth planes immediately surrounding the plane where a hand (or face) was found in the previous frame. Subsequent search in the subset of the intensity data corresponding to these planes produces position of the body part in the current frame. Therefore, intensity data is thresholded for the certain range and depth. Such combined use of input features produces not only speedup due to a significant reduction in a search space, but also increased reliability due to decreased number of false positives that could fall in such space. Instead of processing all pixels (using image subtraction or statistical approaches), this allows us to select only pixels with the certain depth based on the depth of previously detected region of interest.

3 Experimental Results

The following experiments involve application of the algorithm to color and range images sequences of gestures. Triclops color stereo vision system (manufactured by Point Grey Research, Vancouver) is used to capture these sequences. The module connects to a Pentium II MMX PC. Range information is recovered from a correlation-based trinocular stereo algorithm at a speed of up to 30 frames per second.

Two intensity images from a sequence of the speaking person are shown in Figures 1(a-b). More images are not displayed due to space restrictions. Skin detection algorithm is applied to the intensity data from Figure 1(a). Results of skin thresholding following color segmentation are shown in Figure 1(c). Pixels classified as skin are white. Note that, along with the face and hands information, it picks up up parts of other objects - a curtain on the right and a belt. Alternatively, moving objects in Figure 1(a) can be detected by subtracting previous frame (not shown). Whiter areas point to more significant differences between the images.

Instead of applying the same method (color segmentation or subtraction), a proposed approach takes range data into consideration by selecting a depth level(s) where a region of interest was found (Figure 1(e)). Since the motion between two frames is small, the same level can indicate approximate location of the hand in the next frame (red areas in Figure 1(f)). This level along with two closest depth levels (before and after) constitute the search space for current frame (instead of the entire image).

4 Summary and Future Work

We proposed a combination of intensity- and range-based input features. Range data enables localized search for specific features which improves tracking reliability and speed. Proposed approach was success-